Exact Indexing of Dynamic Time Warping

Eamonn Keogh
UC, Riverside

Aarti Bindlish
Bindlish@usc.edu

Main Motive

- Introduces a new technique for exact indexing of Dynamic Time Warping

Outline of Talk

- Why do Time Series Similarity Matching?
- Limitations of Euclidean Distance
- Dynamic Time Warping
- Lower Bounding Dynamic Time Warping
- Indexing Dynamic Time Warping
- Experimental Evaluation
- Conclusions

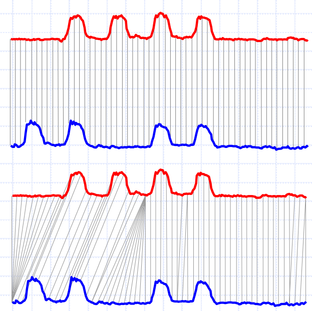
Background

Euclidean Distance

One to one Alignment

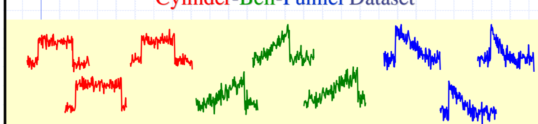
Dynamic Time Warping

Allows elastic shifting of time Axis to accommodate out of phase sequences



Euclidean Vs DTW

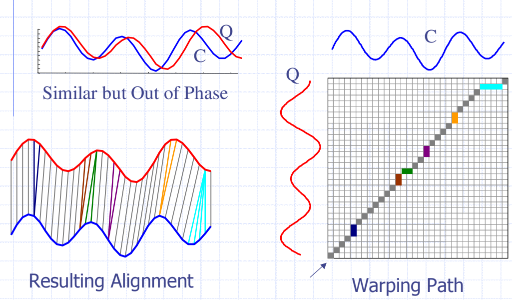
Classification/Clustering Experiment on
Cylinder-Bell-Funnel Dataset



- Dataset of 10 instances of each class
- 1-Nearest Neighbor Algorithm
- Cost : DTW takes 230 times longer

	Mean Error rate
Euclidean	0.2734
DTW	0.0269

DTW Path Construction



Warping Path Constrains

- Boundary Conditions

  W1=(1,1) and wk=(m,n)

- Continuity

  Given Wk=(a,b) then wk-1=(a', b')
  A-a'<=1 and b-b'<=1

- Monotonicity

  Wk=(a,b) then wk-1=(a', b')
  Where a-a'>=0 and b-b'>=0

Optimal Warping Path

- Two Time Series Q and C of length n and m resp.
  Q=q1, q2, q3……qn
  C=c1, c2, c3, …. Cm
- Path with minimum cost – Can be found using Dynamic programming

$$DTW(Q,C) = \min\left\{ \sqrt{\sum_{k=1}^{K} w_k} \Big/ K \right.$$

Cummulative Distance:

$$\gamma(i,j) = d(q_i,c_j) + \min\{ \gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1) \}$$

Why Lower Bound the DTW?

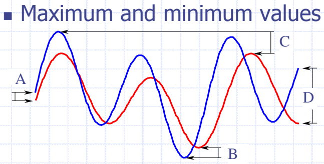
- Search similarity under Euclidean metric is heavily I/O bound
- Search similarity under DTW is also CPU bound
- Pruning sequences which can not possibly be a best match will save considerable amount of time.

Existing Lower Bound Measures

- Lower bounding function introduced by Kim et al. (LB_Kim)
- Second introduced by Yi et al. (LB_Yi)
- Intuition: Cheap & Fast Lower Bound

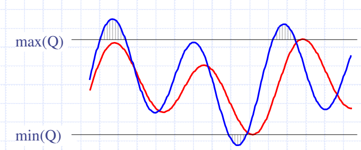
LB_Kim

- Extract 4-tuple feature vector
  - First and last elements of the sequence
  - Maximum and minimum values



- The squared difference between the two sequence's first (A), last (D), minimum (B) and maximum points (C) is returned as the lower bound

LB_Yi



Intuition: All points in one sequence that are larger than the maximum or smaller than minimum must contribute at least the squared difference to the final DTW distance.

Lower bound measure → sum of squared lengths of gray lines

Global Constraints

◆What???
- Places a limit on how far the warping path may stray from the diagonal

◆Why??
- Slightly speed up the DTW distance calculation
- Prevent Pathological warpings

Example

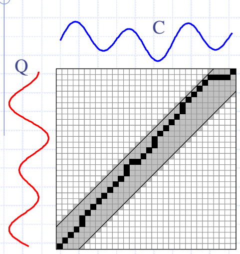
◆A working man can take no more than 3 days leave in a month.

◆He can adjust these leave days in the first following month but not thereafter.

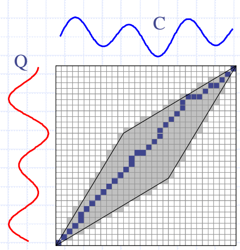
◆It makes sense to allow warpings upto one month. (Warping width of n/12)

◆More drastic warpings should not be allowed.

Warping Windows



Sakoe-Chiba Band

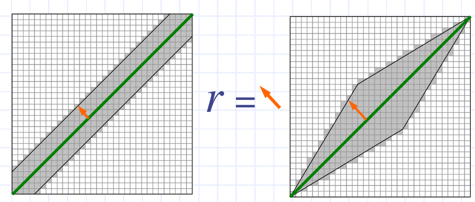


Itakura Parallelogram

A global constraint constrains the indices of the warping path

$w_k = (i,j)_k$ such that $j-r \leq i \leq j+r$

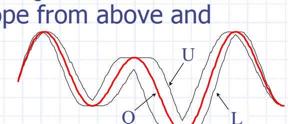
Where r is a term defining allowed range of warping for a given point in a sequence.



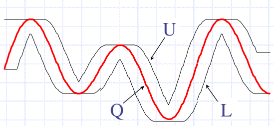
Sakoe-Chiba Band   Itakura Parallelogram

Proposed Lower Bound Approach (LB_Keogh)

◆Original sequence Q is enclosed in a bounding envelope from above and below.



Itakura Parallelogram



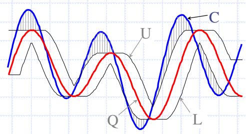
Sakoe-Chiba Band

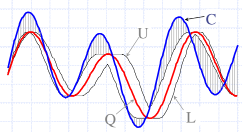
Proposed Lower Bound Approach (LB_Keogh)

$$LB\_Keogh(Q,C) = \sqrt{\sum_{i=1}^{n} \begin{cases} (c_i - U_i)^2 & \text{if } c_i > U_i \\ (c_i - L_i)^2 & \text{if } c_i < L_i \\ 0 & \text{otherwise} \end{cases}}$$

$U_i = \max(q_{i-r} : q_{i+r})$
$L_i = \min(q_{i-r} : q_{i+r})$

LB_Keogh (contd…)





The square sum of distances from every part of the candidate sequence C nto falling within the bounding envelope, to the nearest orthogonal edge of envelope, is returned as lower bound.

The tightness of the lower bound for each technique is proportional to the length of gray lines

A Dimensionality Reduction Technique Piecewise Aggregate Approximation (PAA)

- Time Series C=c1, c2,…..cn (length n)
- N be the dimensionality of space to be indexed (1<=N<=n)
- To reduce Time series from n to N dimensions, the data is divided in to N equal sized frames.

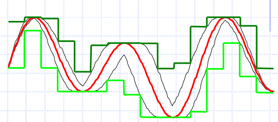
$$\overline{C}_i = \frac{N}{n}\sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} c_j$$

- The mean value of data falling within a frame is calculated. Such a vector of values becomes reduced data representation.

Lower bound for LB_Keogh

$$\hat{U}_i = \max\left(U_{\frac{n}{N}(i-1)+1},\ldots,U_{\frac{n}{N}(i)}\right)$$

$$\hat{L}_i = \min\left(L_{\frac{n}{N}(i-1)+1},\ldots,L_{\frac{n}{N}(i)}\right)$$



$\hat{U}$ and $\hat{L}$ denote PAA which bound U and A resp without intersecting them

Lower bound for LB_Keogh =

$$\sqrt{\sum_{i=1}^{N}\frac{n}{N}\begin{cases}(l_i-\hat{U}_i)^2 & \text{if } l_i>\hat{U}_i\\ (h_i-\hat{L}_i)^2 & \text{if } h_i<\hat{L}_i\\ 0 & \text{otherwise}\end{cases}}$$

LB_PAA

Experimental Philosophy

- Goal
  - To conduct a detailed set of time series indexing experiments
- Specifications
  - Algorithms tested on 32 datasets
  - Uses Sakoe-Chiba Band
  - Reproducible experiments

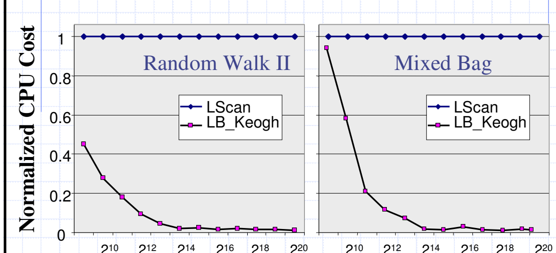
Performance Metric

- Normalized CPU cost

$$\frac{\text{Average CPU time to execute a query using the index}}{\text{Avg CPU time required to perfom linear sequential search}}$$

Normalized cost of linear scan is 1.0

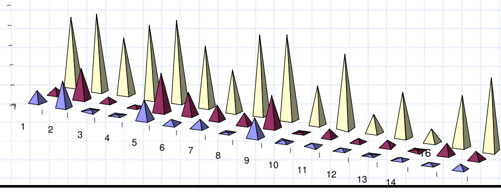
Cost Comparison



Comparison of Tightness of LB Functions

$$T = \frac{\text{Lower Bound Estimate of Dynamic Time Warp Distance}}{\text{True Dynamic Time Warp Distance}}$$

0<=T<=1, Larger the better



Constrains in the approach

◈Only the case where two sequences are of same length

◈Can only index sequences assuming if the warping path is constrained.

Conclusion

- DTW is better distance measure than Euclidean distance.
- Introduced a new lower bounding technique for DTW.
- How to index the new lower bounding technique.
- Demonstrated the utility of this approach with a comprehensive empirical evaluation.

Questions?

Answers?